

All-atom homology model of the *Escherichia coli* 30S ribosomal subunit

Chang-Shung Tung¹, Simpson Joseph² and Kevin Y. Sanbonmatsu¹

¹Theoretical Biology and Biophysics, Los Alamos National Laboratory, Los Alamos, New Mexico 87545, USA. ²Department of Chemistry and Biochemistry, University of California at San Diego, La Jolla, California 92093-0647, USA.

Published online: 16 September, doi:10.1038/nsb841

Understanding the structural basis of ribosomal function requires close comparison between biochemical and structural data. Although a large amount of biochemical data are available for the *Escherichia coli* ribosome, the structure has not been solved to atomic resolution. Using a new RNA homology procedure, we have modeled the all-atom structure of the *E. coli* 30S ribosomal subunit. We find that the tertiary structure of the ribosome core, including the A-, P- and E-sites, is highly conserved. The hypervariable regions in our structure, which differ from the structure of the 30S ribosomal subunit from *Thermus thermophilus*, are consistent with the cryo-EM map of the *E. coli* ribosome.

The ribosome is a dynamic molecular machine that manufactures protein according to the genetic information residing in nucleic acid^{1,2}. Because several large conformational changes occur during protein synthesis, it is necessary to relate the atomic resolution structures of the ribosome in its various functional states to functional information derived from biochemical experiments. Because the majority of biochemical protein synthesis experiments have been performed using ribosomes from *Escherichia coli*^{3–4}, a detailed structure of the *E. coli* ribosome would be required to put all the biochemical information into its appropriate structural context. However, the high-resolution structure of the *E. coli* ribosome has not yet been determined experimentally.

Homology modeling has become a commonly used technique to predict protein structure in recent years⁵. Although RNA is closely related to DNA chemically, it can fold into a more complicated tertiary structure⁶ that is often observed in proteins. The complex tertiary structures of RNA molecules present a challenge for developing RNA homology modeling methods analogous to those developed for proteins. Before 2000, the limited availability of high-resolution RNA structures made modeling large RNA three-dimensional structures using exclusively computational approaches difficult. Often, the three-dimensional structure of RNA was modeled using a knowledge-based approach in combination with the application of constraints derived from experimental observations^{7–10}. The high resolution structures of both the 30S ribosomal subunit for *Thermus thermophilus*^{2,11} and the 50S ribosomal subunit for *Haloarcula marismortui*¹² and *Deinococcus radiodurans*¹³ have significantly extended the knowledge base of RNA three-dimensional structure. With the extended knowledge base, modeling RNA structures using a homology modeling approach is now possible.

Homology modeling is based on the principle that molecules with similar sequences and similar functions should fold into similar three-dimensional structures. In practice, a template molecule can be identified with a known structure and sequence similarity to the target molecule. Here, we present a homology model of the atomic structure of the *E. coli* 30S ribosomal subunit using the crystal structure of the *T. thermophilus* 30S ribosomal subunit¹ as the template (PDB entry 1FJF).

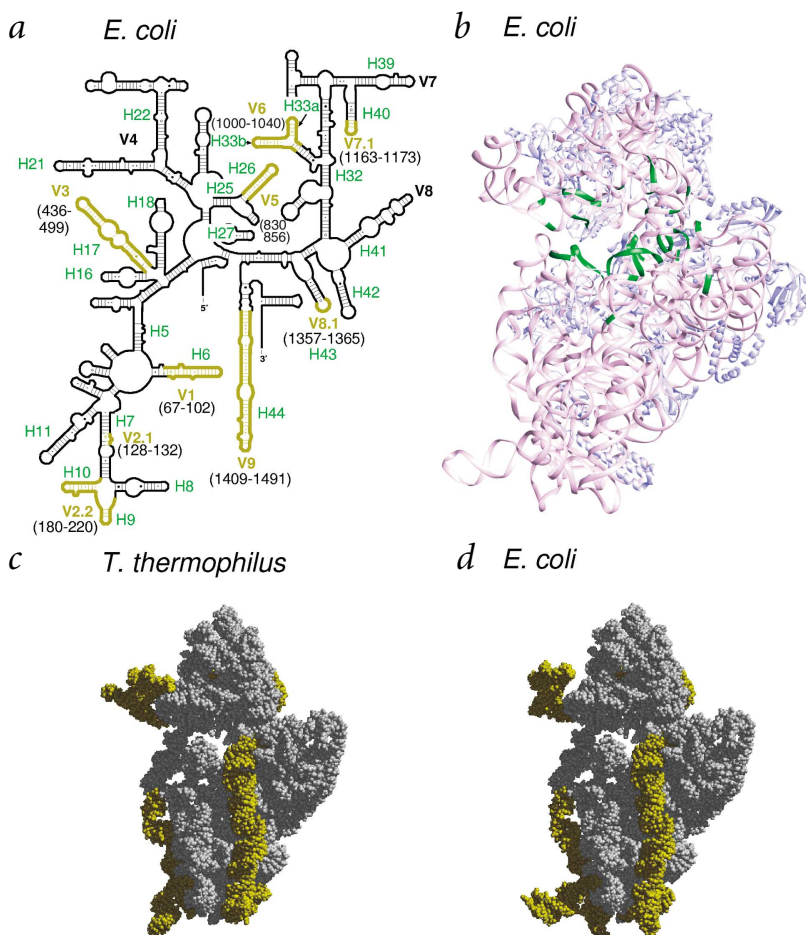


Fig. 1 Secondary and tertiary structures of the 16S rRNA. **a**, Variable regions of *E. coli* 16S rRNA that contain deletions/insertions relative to *T. thermophilus* are shown in yellow. Variable regions that do not contain deletions/insertions are shown in black. Helices are labeled according to the Gutell numbering scheme. **b**, The minimized structure of the *E. coli* 30S ribosomal subunit. RNA is shown in pink, and protein is shown in purple. The universally conserved bases not involved in tertiary contacts are shown in green. These bases line the surface of the functional sites of the molecule. Structure of the **c**, *T. thermophilus* 16S rRNA and **d**, *E. coli* 16S rRNA. The variable regions, which involve motif modeling, are shown in yellow. All these regions are on the surface of the molecule.

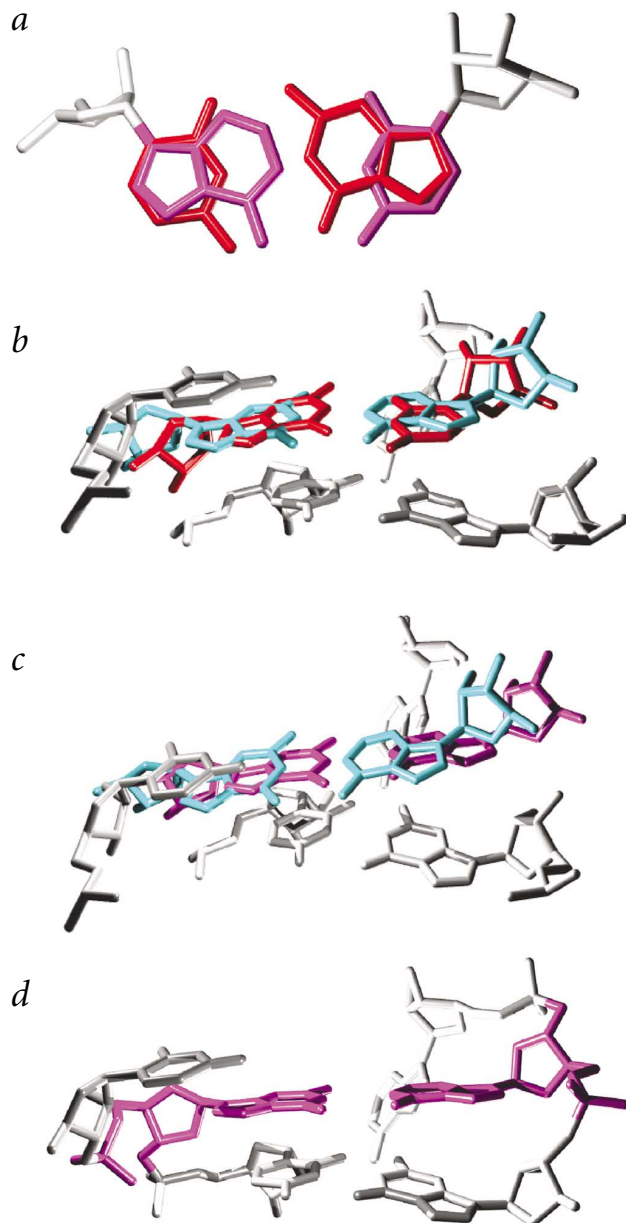


Fig. 2 Homology model of a base pair in a helix. **a**, The structure of a template C-G pair (red) is substituted to a target A-U pair (magenta). **b**, The superposition of the template U-G (red) and the intermediate target A-G (blue) base pair. **c**, The reoriented structure of the final target A-G base pair (magenta) and the intermediate target A-G base pair (blue). **d**, The structure of the final target nucleotides (magenta) with the neighboring nucleotides.

base pairs that are significantly different in size from those of Watson-Crick base pairs, structural adjustments are required. We have shown the base substitution from U129•G232 in *T. thermophilus* 16S rRNA to A129•G232 in *E. coli* as an example (Fig. 2b–d). The first step of the procedure (Fig. 2b) involves the superposition of the mean base pair planes and the base pair centers between an intermediate target base pair (blue) and the template base pair (red). The ribose conformations and the glycosidic angles for the target nucleotides are kept identical to those of the template nucleotides. The line connecting the C1' atoms of the two nucleotides in the intermediate target base pair is set to be parallel to that of the template base pair. This particular consideration makes the helix twist angle of the target base pair similar to that of the template base pair and provides a good estimate for the phosphate connections. In the case where a base pair substitution results in phosphate connections larger than the maximum distance (4.5 Å) for a phosphate group to bridge across, the two nucleosides forming the intermediate target base pair are subjected to small rotation and translation to search for orientations that allow phosphate connections. Proper van der Waals contacts with the neighboring nucleosides are maintained during the search process. The orientation of the nucleosides before (blue, intermediate target) and after (magenta, final target) the search are shown in Fig. 2c. The last step of the procedure involves the construction of the phosphate group using a phosphate-modeling algorithm developed in our laboratory¹⁸. The structure of the final target nucleotides with the neighboring nucleotides is shown in Fig. 2d. In addition to base pair A129•G232, the structures of three other base pairs (from G606•A632 to G606•U632, from A611•G629 to C611•A629 and from A684•A706 to U684•A706) are modeled using this procedure.

In regions where the alignment involves deletions or insertions, a motif-modeling approach is used. The approach essentially divides the RNA region to be modeled into fragments, each corresponding to a motif with known structure. These motifs can be assembled by aligning the bases in the overlapping regions. Because the structures of the overlapping regions in the motifs used are not identical, adjustments of the phosphate structure¹⁸ at the junctions are required during the assembly process. The secondary structure of the spur region of the *E. coli* 16S rRNA indicates that the spur can be interpreted as a large hairpin (Fig. 3a) containing three bulges and one tetraloop. This large hairpin structure is divided into five fragments (Fig. 3b). These fragments have overlapping regions consisting of at least one base pair. To model the one-nucleotide bulge, we search for such a motif in the template crystal structure¹. There are four one-nucleotide bulge motifs that involve a purine base (two with guanines and two with adenines) at the bulge. Among these four motifs, three have the bulge base in the stack, whereas one has the bulge base-flipped out of the helix. The motif with the bulge base flipped out has a guanine at the bulge and two G•U base pairs flanking the bulge base. The structures of the base pairs in the remaining three one-purine bulge motifs are very similar. The main difference between the structures of

Sequence alignment of 16S rRNAs

The alignment reveals a high sequence identity (75%) between *E. coli* and *T. thermophilus* 16S rRNA sequences. The two sequences can be aligned without insertions or deletions, apart from nine short stretches of RNA that all reside in the variable regions (V1–V9) defined by Neefs *et al.*¹⁴ and are used for species identification. Regions 2.1 and 2.2 are subregions of V2. Regions 7.1 and 8.1 are slightly displaced from V7 and V8, as shown from the secondary structure of the 16S rRNA^{15–16} (Fig. 1a).

Modeling of the *E. coli* 30S ribosome

Two types of approaches are implemented to model *E. coli* 16S rRNA structure. For regions where the sequences are aligned without deletion or insertion, direct base substitution is applied. Because the C1'...C1' distance is essentially the same (~10.2 Å) in all Watson-Crick base pairs¹⁷, this simple procedure preserves the base pairing and the double helical structure while substituting the bases (Fig. 2a). For non-Watson-Crick

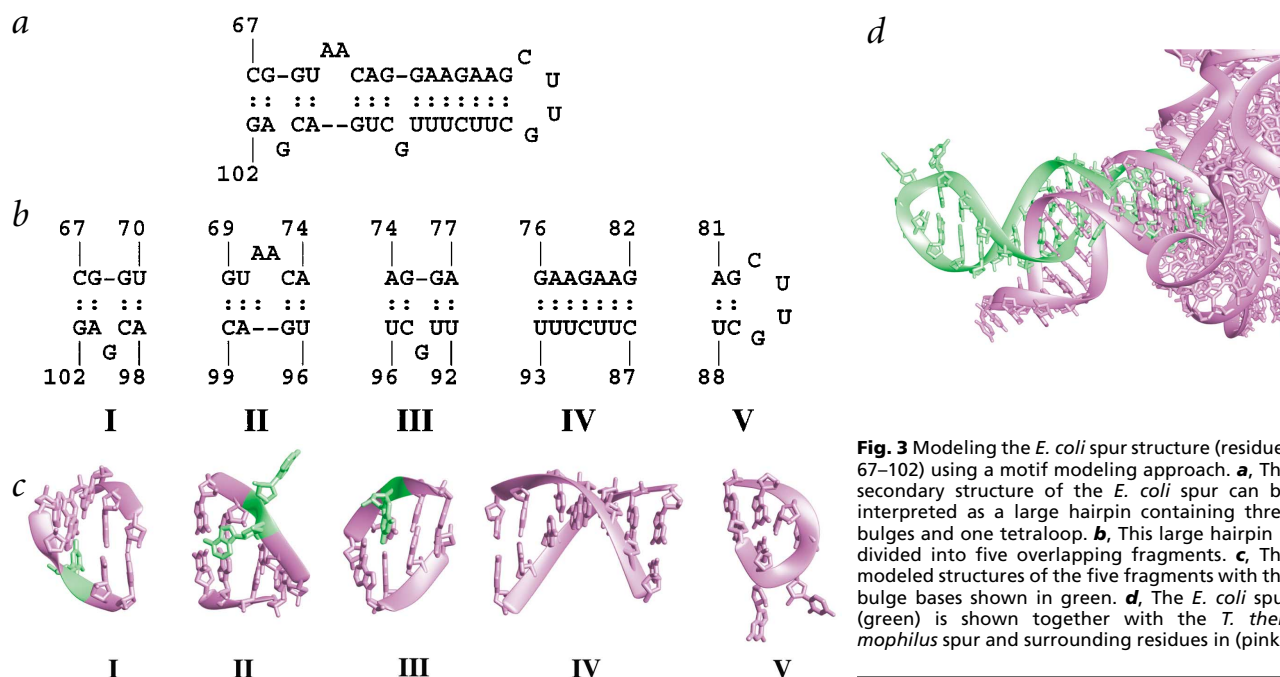


Fig. 3 Modeling the *E. coli* spur structure (residues 67–102) using a motif modeling approach. **a**, The secondary structure of the *E. coli* spur can be interpreted as a large hairpin containing three bulges and one tetraloop. **b**, This large hairpin is divided into five overlapping fragments. **c**, The modeled structures of the five fragments with the bulge bases shown in green. **d**, The *E. coli* spur (green) is shown together with the *T. thermophilus* spur and surrounding residues (in pink).

the A-bulge and the G-bulge is that the A-bulge base is planar to its 5' base pair, whereas the G-bulge base is planar to its 3' base pair. This planar arrangement of the bulge base to a flanking base pair is due to a hydrogen bond formed between the bulge base and a G on the opposite strand in the flanking base pair. The two fragments with one nucleotide bulge (green, Fig. 3c) to be modeled (fragments I and III) have a G base at the bulge and a G base at the opposite strand of the base pair 3' to the bulge. This is consistent with the G-bulge motif (residues 593–597 and 643–646) in the template structure. Therefore, we choose to use this G-bulge motif as a template to model the structures for fragments I and III.

We use a similar procedure to identify the GA-bulge motif (residues 128–132 and 230–233) that is used to model fragment II. There is one C-U-U-G tetra-loop hairpin motif (residues 190A–190H based on 1FJF numbering) in the template structure, which we use to model fragment V. We use the seven base pairs in front of the hairpin loop of the corresponding hairpin in the template structure as the template to model fragment IV (Fig. 3c). Overall, the absence of serious steric clashes in the structure derived from the homology modeling (Fig. 3d) supports the idea that the evolution process has preserved the three-dimensional structures, as well as tertiary contacts, in the core and functional regions.

The sequence identities between the two sets of ribosomal protein sequences are 30–73% (Table 1). In general, when the sequence identity of target and the template proteins is $\geq 30\%$, homology modeling can be applied⁵. We modeled the structures of the *E. coli* ribosomal proteins S2–S20 using the homology modeling approach of Tung¹⁹. In this modeling approach, the side chain atoms are attached to the main chain with torsional angles maintained close to those in the template structure¹⁹. This approach has the advantage of allowing both the target and the template residues to use the same, often limited, space available to the side chains. We find this consideration particularly useful when modeling structures of protein in protein–RNA complexes. The alignment of the two S2 sequences is poor at the C-terminal region (the last 26 residues). To avoid uncertainty in the

homology model, we have truncated the *E. coli* S2 model at residue 214. Protein S14 is significantly larger in *E. coli* than in *T. thermophilus*. The *T. thermophilus* S14 sequence is aligned to the C-terminal part of the *E. coli* S14 sequence, with the extra *E. coli* S14 residues located in the N-terminus. To improve the stereochemistry of the modeled proteins, each of the modeled protein structures is subjected to energy minimization (1,000 steps) using the AMBER²⁰ suite of molecular dynamics and minimization routines. The structures of 16 of the energy-minimized ribosomal proteins (S2–S7, S9–S16, S18 and S20) are

Table 1 Properties of *E. coli* 30S ribosomal proteins

	Residues	% identity ¹	Δ_{cm} (Å) ¹	Interface surface area (Å ²) ²
S2	6–214	46	0.09	1,895
S3	2–207	52	0.50	3,353
S4	2–206	49	0.78	6,001
S5	10–159	46	0.15	3,837
S6	1–100	30	0.60	1,131
S7	3–156	53	0.29	3,593
S8	2–129	52	2.05	3,691
S9	4–130	54	0.26	5,803
S10	5–102	54	0.38	3,629
S11	13–129	54	0.28	3,725
S12	2–123	73	0.34	6,593
S13	2–118	57	1.20	4,574
S14	42–101	35	0.40	3,870
S15	2–89	59	0.33	3,527
S16	1–82	34	0.46	5,708
S17	4–84	47	0.33	5,991
S18	5–75	41	1.48	1,744
S19	3–81	67	0.18	3,199
S20	3–87	33	0.11	6,140

¹% identity and Δ_{cm} were calculated between proteins in *E. coli* and *T. thermophilus*.

²Interface surface area calculated between the proteins and the 16S rRNA in 30S ribosomal subunit.

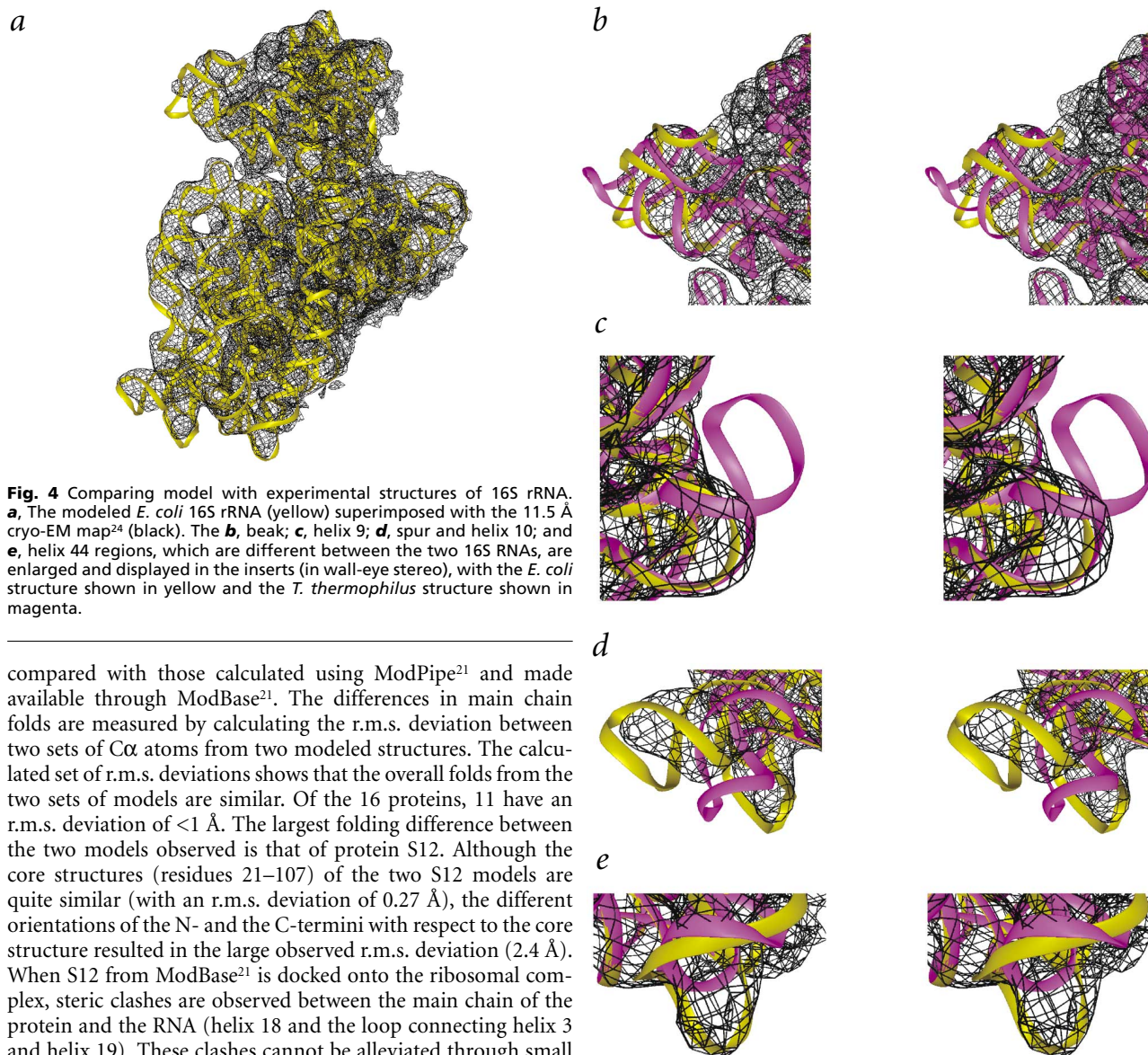


Fig. 4 Comparing model with experimental structures of 16S rRNA. **a**, The modeled *E. coli* 16S rRNA (yellow) superimposed with the 11.5 Å cryo-EM map²⁴ (black). The **b**, beak; **c**, helix 9; **d**, spur and helix 10; and **e**, helix 44 regions, which are different between the two 16S RNAs, are enlarged and displayed in the inserts (in wall-eye stereo), with the *E. coli* structure shown in yellow and the *T. thermophilus* structure shown in magenta.

compared with those calculated using ModPipe²¹ and made available through ModBase²¹. The differences in main chain folds are measured by calculating the r.m.s. deviation between two sets of C α atoms from two modeled structures. The calculated set of r.m.s. deviations shows that the overall folds from the two sets of models are similar. Of the 16 proteins, 11 have an r.m.s. deviation of <1 Å. The largest folding difference between the two models observed is that of protein S12. Although the core structures (residues 21–107) of the two S12 models are quite similar (with an r.m.s. deviation of 0.27 Å), the different orientations of the N- and the C-termini with respect to the core structure resulted in the large observed r.m.s. deviation (2.4 Å). When S12 from ModBase²¹ is docked onto the ribosomal complex, steric clashes are observed between the main chain of the protein and the RNA (helix 18 and the loop connecting helix 3 and helix 19). These clashes cannot be alleviated through small rigid-body translation and rotation. Similar clashes were observed when S9, S10 and S14 from ModBase²¹ were docked into the 30S ribosomal complex. No such clashes were observed from our protein models. Therefore, the best strategy in the homology modeling of these ribosomal proteins is to preserve their folds from the template structures.

The model of the 30S ribosomal subunit is obtained by combining structures of the 16S rRNA and the ribosomal proteins (S2–S20). The homology modeling approach allows the preservation of favorable interactions between the RNA and the proteins in the ribosomal subunit. To ensure that good stereochemistry exists in the ribosomal complex — that is, the absence of improper van der Waals contacts between the RNA and the proteins — the structure of the complex is subjected to additional energy minimization (1,000 cycles) using AMBER²⁰ (Fig. 1b).

Structural comparison between the two 16S rRNAs

The regions of deletion/insertion in 16S rRNA of *E. coli* and *T. thermophilus*, as viewed from the interface side (yellow, Fig. 1c,d), all lie in the variable regions as defined in Neefs *et al.*¹⁴.

With the exception of helix 44, all these regions are on the exterior of the molecule, away from the 30S ribosomal A- and P-sites. The 5' domain of the 16S rRNA contains V1, V2.1, V2.2 and V3. There are noticeable differences between the two spur structures (V1 and H6) in the 5' domain. For the *E. coli* 16S rRNA, the spur is longer and pointed horizontally outward as compared with the spur of the *Thermus thermophilus* 16S rRNA. The change in the orientation of the *E. coli* spur relative to *T. thermophilus* is due to the added three bulges that result in the change of the helix direction. Two other differences between the two structures occur in helix 9 and helix 10. Helix 9 is longer and helix 10 is shorter in *T. thermophilus* 16S rRNA relative to those of the *E. coli*. In the central domain (platform), the two rRNA sequences can be aligned without insertions or deletions, with the exception of helix 26. This helix is slightly shorter in *T. thermophilus* as compared with that in *E. coli*. The 3' major domain contains V6, V7.1 and V8.1. V6 includes helices 33a and 33b, commonly referred to as the 'beak'. The difference in length in these helices between *E. coli* and *T. thermophilus* results in a differently shaped beak for *E. coli*. The 3' minor domain (penultimate stem) consists of helix

44, which contains the decoding region and bases that interact with the large subunit²². Except for the difference in length, the overall structure of helix 44 is conserved between *E. coli* and *T. thermophilus* despite difference in sequence.

Model and cryo-EM map of the *E. coli* 30S subunit

Using SUPCOMB²³, the *E. coli* 16S rRNA was docked into the *E. coli* 70S ribosome cryo-EM map²⁴ (Fig. 4a). In all five regions that show significant structural differences — spur, beak, helix 9, helix 10 and helix 44 (Figs 4b–e) — the modeled structures of the docked 16S *E. coli* rRNA (yellow) fit the cryo-EM map (black) significantly better than those in the *T. thermophilus* rRNA (magenta). The normalized spatial discrepancy²³, calculated by SUPCOMB²³, is 1.1 for *E. coli* and 1.15 for *T. thermophilus*, consistent with the observation that the regions of difference represent ~5% of the 16S rRNA. We emphasize that the comparison between our *E. coli* homology model and the *E. coli* cryo-EM map was obtained solely *via* rigid body docking. No modifications to the structure were made to fit the cryo-EM map.

Proteins and rRNA interactions in *E. coli* 30S ribosome

Overall, the two sets of ribosomal proteins (from *T. thermophilus* and *E. coli*) interact with their corresponding 16S rRNA in a similar manner. This can be seen in the differences in center-of-mass (Δ_{cm}) between the corresponding proteins with respect to the superimposed 30S ribosomal subunit complex structures. The Δ_{cm} values for the set of proteins are small, ranging from 0.09 to 2.05 Å (Table 1). Except for S6, all proteins of the 30S subunit are basic proteins, with the number of excess basic residues ($N_{excess} = N_{basic} - N_{acid}$) ranging from 5 (slightly basic) to 24 (highly basic). Similar to histone proteins in the nucleosome, these basic proteins can serve as stabilizing factors in the folding of the highly negatively charged RNA molecule. To study the interactions between the RNA and proteins in the complex, interface surface areas were calculated using Naccess²⁵. The calculated interface surface areas (Table 1) range from 1,131 Å² (between S6 and the rRNA) to 6,593 Å² (between S12 and the rRNA), with an average value of 4,105 Å². These values are consistent with the 1,120–5,800 Å² interface surface areas as observed from 75 X-ray structures of protein–nucleic acid complexes²⁶. Out of the 19 ribosomal proteins, the most basic protein is S12. Besides having the largest interface surface area, the sequence of S12 is also the most conserved between the two sets of proteins (73% identity). S12 is of significant functional importance and is the only protein that interacts with the decoding site bases (in particular A1492) of 16S rRNA. The functional significance of S12 is also demonstrated by streptomycin, which induces misreading by binding to S12 (ref. 11). Streptomycin-resistant and -dependent strains often involve mutations in S12. Finally, S12 is also significant as the binding site of initiation factor 1 (ref. 27).

Sequence and structure conservation in 16S rRNA

Because of structural and functional requirements, the 16S rRNA sequences from different organisms are highly conserved. From a folding point of view, duplex structures are similar regardless of sequence. If functional conservation requires folding conservation, then one can propose that the sequence conservation should be higher in the loop regions of the rRNA. To test this hypothesis, we checked the sequence conservation between *E. coli* and *T. thermophilus* 16S rRNA sequences. The two aligned sequences are divided into helix and loop regions. The sequence identities between the two sequences are 68% for

the helix and 82% for the loop regions. Indeed, the loop regions have higher sequence conservation than the helix regions. Sequence conservation studies of the 16S rRNA across species²⁸ have identified 213 universally conserved bases, which are distributed throughout the 16S rRNA tertiary structure. Out of these universally conserved residues, 148 are involved in either intramolecular tertiary contacts (between residues that are distant in sequence) or intermolecular contacts (between 16S rRNA bases and ribosome protein residues). These contacts are important in stabilizing the tertiary fold of the RNA, as well as the complex formation of the ribosome. The remaining 65 universally conserved residues (green residues, Fig. 1b) line the surface of the functional sites (A-site, P-site, E-site and inter-subunit bridges), delineating the path of mRNA and tRNAs through the ribosome. These bases are either near the path of mRNA²⁹ or tRNA²², or interact with the 23S rRNA. This observation is consistent with the notion that the conservation in structures of functional sites results in the conservation of sequences at these sites.

Methods

RNA sequences and alignment. The sequence of the *E. coli* 16S rRNA was obtained from GenBank³⁰ with the accession number J01695³¹, and the sequence of the *T. thermophilus* 16S rRNA is from the crystal structure of PDB³² entry 1FJF¹. The sequence lengths of the two 16S rRNAs are slightly different (1,515 and 1,542 nucleotides for *T. thermophilus* and *E. coli*, respectively). The alignment of the two nucleotide sequences was obtained using ALIGN³³ and confirmed using BLAST 2 Sequences³⁴. The secondary structural information was used for the alignment of the two sequences in the regions involving insertion/deletion. The secondary structures of the two 16S rRNA can be obtained from the Comparative RNA Web Site (<http://www.rna.icmb.utexas.edu>).

Protein sequences and alignment. The sequences of the *E. coli* 30S ribosomal proteins have been published³⁵. In this study, we used the sequences determined by Makino *et al.*³⁶ and obtained from the GenBank Sequence Database³⁰. The alignments between these protein sequences from *T. thermophilus* and *E. coli* are obtained using ALIGN³³ and confirmed using BLAST 2 Sequences³⁴.

Docking of the 16S rRNA model into the cryo-EM map. The solution structure of the *E. coli* ribosome is derived from the 11.5 Å resolution cryo-EM map of the *E. coli* 70S ribosome²⁴. To obtain the cryo-EM map for the small subunit, the grid points with high density that correspond to the large subunit are removed manually. Docking of the modeled structure into the cryo-EM map is done using SUPCOMB²³. SUPCOMB aligns two three-dimensional objects by minimizing a normalized spatial discrepancy (NSD)²³. For the purpose of alignment, those points in the cryo-EM map that exceed a threshold (>110) are selected to represent the cryo-EM model. Only the phosphorus atoms in the 16S rRNA are used for alignment.

Coordinates. The coordinates of the homology model have been deposited in the Protein Data Bank (accession code 1M5G).

Acknowledgements

This work was performed under the auspices of the Department of Energy (DOE) under contract to the University of California. C.S.T. and K.Y.S. are supported by LANL/LDRD funding. S.J. is supported by both NSF and NIH grants. The cryo-EM density map was kindly provided by J. Frank of HHMI.

Competing interests statement

The authors declare that they have no competing financial interests.

Correspondence should be addressed to C.-S.T. email: ct@lanl.gov

Received 28 May, 2002; accepted 15 August, 2002.

1. Wimberly, B.T. et al. *Nature* **407**, 327–339 (2000).
2. Schlutzen, F. et al. *Cell* **102**, 615–623 (2000).
3. Green, R. & Noller, H.F. *Annu. Rev. Biochem.* **66**, 679–716 (1997).
4. O'Connor, M. in *The Ribosome: Structure, Function, Antibiotics, & Cellular Interactions* (eds Garrett, R.A. et al.) 217–227 (ASM Press, Washington, D.C.; 2000).
5. Sanchez, R. & Sali, A. *Curr. Opin. Struct. Biol.* **7**, 206–214 (1997).
6. Westhof, E. *Theochem* **105**, 203–210 (1993).
7. Westhof, E. & Altman, S. *Proc. Natl. Acad. Sci. USA* **91**, 5133–5137 (1994).
8. Stern, S., Weiser, B. & Noller, H.F. *J. Mol. Biol.* **204**, 447–481 (1988).
9. Malhotra, A. & Harvey, S.C. *J. Mol. Biol.* **240**, 308–340 (1994).
10. Mueller, F. & Brimacombe, R. *J. Mol. Biol.* **271**, 524–544 (1997).
11. Carter, A.P. et al. *Nature* **407**, 340–348 (2000).
12. Ban, N., Nissen, P., Hansen, J., Moore, P.B. & Steitz, T.A. *Science* **289**, 905–920 (2000).
13. Harms, J. et al. *Cell* **107**, 679–688 (2001).
14. Neefs, J., Van de Peer, Y., De Rijk, P., Chapelle, S. & De Wachter, R. *Nucleic Acids Res.* **21**, 3025–3049 (1993).
15. Guetell, R.R. *Nucleic Acids Res.* **22**, 3502–3507 (1994).
16. Guetell, R.R. in *Ribosomal RNA. Structure, Evolution, Processing, and Function in Protein Biosynthesis* (eds Dahlberg, A. & Zimmerman, B.) 111–128 (CRC Press, Boca Raton; 1996).
17. Watson, J.D. & Crick, F.H.C. *Nature* **171**, 737–738 (1953).
18. Tung, C.S. in *Computation of Biomolecular Structures* (eds Soumpasis, D.M. & Jobin, T.M.) 87–97 (Springer-Verlag, New York; 1993).
19. Tung, C.S. *J. Biomol. Struct. Dyn.* **17**, 347–354 (1999).
20. Weiner, S.J., Kollman, P.A., Nguyen, D.T. & Case, D.A. *J. Comput. Chem.* **7**, 230–252 (1986).
21. Pieper, U., Eswar, N., Stuart, A.C., Llyin, V.A. & Sali, A. *Nucleic Acids Res.* **30**, 255–259 (2002).
22. Yusupov, M.M. et al. *Science* **292**, 883–896 (2001).
23. Kozin, M.B. & Svergun, D.I. *J. Appl. Crystallogr.* **34**, 33–41 (2001).
24. Gabashvili, I.S. et al. *Cell* **100**: 537–549 (2000).
25. Hubbard, S.J. & Thornton, J.M. Naccess. (Department of Biochemistry and Molecular Biology, University College, London; 1993).
26. Nadassy, K., Wodak, S.J. & Janin, J. *Biochemistry* **38**, 1999–2017 (1999).
27. Carter, A.P. et al. *Science* **291**, 498–501 (2001).
28. Noller, H.F. In *The RNA World* (eds Gesteland, R.F., Cech, T.R. & Atkins, J.F.) 197–219 (Cold Spring Harbor Laboratory Press, New York 1999).
29. Yusupova, G.Z., Yusupov, M.M., Cate, J.H.D. & Noller, H.F. *Cell* **106**, 233–241 (2001).
30. Benson, D.A. et al. *Nucleic Acids Res.* **28**, 15–18 (2000).
31. Brosius, J., Palmer, M.L., Kennedy, P.J. & Noller, H.F. *Proc. Natl. Acad. Sci. USA* **75**, 4801–4805 (1978).
32. Berman, H.M. et al. *Nucleic Acids Res.* **28**, 235–242 (2000).
33. Person, W.R., Wood, T., Zhang, Z. & Miller, W. *Genomics* **46**, 24–36 (1997).
34. Tatusova, T.A. & Madden, T.L. *FEMS Microbiol. Lett.* **174**, 247–250 (1999).
35. Wittmann, H.G. *Annu. Rev. Biochem.* **51**, 155–183 (1982).
36. Makino, K. et al. *Genes Genet. Syst.* **74**, 227–239 (1999).